A note on nonparametric estimation of bivariate tail dependence

Axel Bücher*
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

July 16, 2012

Abstract

Nonparametric estimation of tail dependence can be based on a standardization of the marginals if their cumulative distribution functions are known. In this paper it is shown to be asymptotically more efficient if the additional knowledge of the marginals is ignored and estimators are based on ranks. The discrepancy between the two estimators is shown to be substantial for the popular Clayton model. A brief simulation study indicates that the asymptotic conclusions transfer to finite samples.

Keywords and Phrases: asymptotic variance, nonparametric estimation, rank-based inference, tail copula, tail dependence

AMS Subject Classification: Primary 62G32 ; secondary 62G05

1 Introduction

Suppose $(X,Y),(X_1,Y_1),\dots,(X_n,Y_n)$ are independent random vectors with joint cdf H and continuous marginal cdfs F and G. The cdf C of $(U,V)=(F(X),G(Y))$ is called the copula of (X,Y) and satisfies the relationship $H(x,y)=C(F(x),G(y))$. The lower and upper tail copulas of (X,Y) (or the lower and upper functions of tail dependence) are defined as the following directional derivatives of the copula C and its associated survival copula $\bar C(u,v)=u+v-1+C(1-u,1-v)$ at the point $(0,0)$:

$$\Lambda_L(x,y)=\lim_{t\to 0}\Pr(F(X)\le tx \mid G(Y)\le ty)=\lim_{t\to 0}\frac{C(tx,ty)}{t},$$
$$\Lambda_U(x,y)=\lim_{t\to 0}\Pr(F(X)\ge 1-tx \mid G(Y)\ge 1-ty)=\lim_{t\to 0}\frac{\bar C(tx,ty)}{t},$$

where $(x,y)\in[0,\infty]^2\setminus\{(\infty,\infty)\}$. Note that $\bar C$ is the joint cdf of the vector $(1-U,1-V)$. Tail copulas and variants thereof characterize extremal dependence of the vector (X,Y), see de Haan and Ferreira (2006). One of the variants of Λ_U is given by the stable tail dependence function $l(x,y)=x+y-\Lambda_U(x,y)$. The restriction of Λ_U to the unit sphere with respect to the $\|\cdot\|_1$-norm, i.e., the function $A(t)=\Lambda_U(1-t,t)$ for $t\in[0,1]$, is called Pickands dependence function, see Pickands (1981). Since tail copulas are homogeneous in the sense that $\Lambda_U(sx,sy)=s\Lambda_U(x,y)$ for all $s>0$, the Pickands dependence function and the upper

*E-mail: axel.buecher@ruhr-uni-bochum.de, Tel.: (+49) 234 / 32 23284, Fax: (+49) 234 / 32 14559

tail copula are one-to-one. Similar remarks can be made for lower tail copulas, even though their variants are less prominent. Of course, the popular coefficients of tail dependence are included in the concept of tail copulas and are given by $\lambda_L = \Lambda_L(1,1)$ and $\lambda_U = \Lambda_U(1,1)$.
Assuming that (U, V) is in the domain of attraction of a bivariate extreme value distribution, the estimation of tail copulas has been addressed by Huang (1992). The underlying idea of her estimator (and of several variants) can be summarized as follows. For the sake of brevity we restrict ourselves to lower tail copulas and we begin by supposing that the marginal distributions are known to the statistician. In that case, a natural estimator for C is given by the empirical distribution function of the standardized sample $(U_1, V_1), \ldots (U_n, V_n)$, i.e., $\tilde{C}_n(u,v) = n^{-1}\sum_{i=1}^n \mathbb{I}\{U_i \le u, V_i \le v\}$. A promising estimator for Λ_L is then given by

$$\tilde{\Lambda}_L(x,y) = \frac{n}{k}\tilde{C}_n\left(\frac{kx}{n}, \frac{ky}{n}\right) = \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{F(X_i) \le kx/n, G(Y_i) \le ky/n\},$$

where $k = k_n$ is a constant that needs to be chosen by the statistician. Relaxing the assumption of having knowledge of the marginal distributions we must replace $\tilde{C}_n(u,v)$ by the empirical copula $\hat{C}_n$. With the *pseudo observations* $\hat{U}_i = F_n(X_i)$ and $\hat{V}_i = G_n(Y_i)$, where F_n and G_n denote the marginal empirical distribution functions, the empirical copula is defined as $\hat{C}_n(u,v) = n^{-1}\sum_{i=1}^n \mathbb{I}\{\hat{U}_i \le u, \hat{V}_i \le v\}$. Hence, we can define

$$\hat{\Lambda}_L(x,y) = \frac{n}{k}\hat{C}_n\left(\frac{kx}{n}, \frac{ky}{n}\right) = \frac{1}{k}\sum_{i=1}^n \mathbb{I}\{\hat{F}(X_i) \le kx/n, \hat{G}(Y_i) \le ky/n\}.$$

The asymptotics of these estimators (or slight variants thereof) have been investigated in Huang (1992); Drees and Huang (1998); Schmidt and Stadtmüller (2006); Einmahl et al. (2012); Bücher and Dette (2011), among others. In order to control the bias of the estimators one needs to assume a second order condition on the speed of convergence in the defining relation for Λ. Suppose that

$$|\Lambda_L(x,y) - tC(x/t, y/t)| = O(B(t))$$

for $t \to \infty$ locally uniformly in (x,y), where $B : [0,\infty) \to [0,\infty)$ satisfies $\lim_{t\to\infty} B(t) = 0$. Then for $k \to \infty, k = o(n)$ such that $\sqrt{k}B(n/k) = o(1)$ it is well known that

$$\sqrt{k}(\tilde{\Lambda}_L(x,y) - \Lambda_L(x,y)) \rightsquigarrow \mathbb{B}_\Lambda(x,y)$$

in the space $(\ell^\infty([0,T]^2), \|\cdot\|_\infty)$ for each $T > 0$, see, e.g., Schmidt and Stadtmüller (2006). Here, $\mathbb{B}_{\Lambda_L}$ denotes a tight centered Gaussian field with covariance structure given by

$$\mathrm{Cov}\{\mathbb{B}_\Lambda(x,y), \mathbb{B}_\Lambda(s,t)\} = \Lambda_L(x \wedge s, y \wedge t).$$

Note that the definition of $\mathbb{B}_\Lambda$ can be extended to the set $[0,\infty]^2 \setminus \{(\infty,\infty)\}$. Using the margin-free estimator $\hat{\Lambda}_L$, and assuming that the partial derivatives $\dot{\Lambda}_{L,1} = \frac{\partial}{\partial x}\Lambda_L$ and $\dot{\Lambda}_{L,2} = \frac{\partial}{\partial y}\Lambda_L$ exist and are continuous on $(0,\infty)^2$, it holds

$$\sqrt{k}(\hat{\Lambda}_L(x,y) - \Lambda_L(x,y)) \rightsquigarrow \mathbb{G}_\Lambda(x,y),$$

where the process $\mathbb{G}_\Lambda$ can be expressed as

$$\mathbb{G}_\Lambda(x,y) = \mathbb{B}_\Lambda(x,y) - \dot{\Lambda}_{L,1}(x,y)\mathbb{B}_\Lambda(x,\infty) - \dot{\Lambda}_{L,2}(x,y)\mathbb{B}_\Lambda(\infty,y),$$

see Einmahl et al. (2012); Bücher and Dette (2011). In the case of unknown marginal distributions, a statistician being interested in tail dependence has no choice between the estimators $\tilde{\Lambda}_L$ and $\hat{\Lambda}_L$ and has to rely on $\hat{\Lambda}_L$. The question of interest of this note deals with case of having knowledge of the marginal distributions. Then we are confronted with the question of which estimator to prefer, and observing that $\tilde{\Lambda}_L$ exploits additional knowledge and has the somewhat easier limiting distribution might suggest to use this estimator. We are going to show that this conclusion is misleading: even though $\hat{\Lambda}_L$ is discarding what appears to be pertinent information, this estimator is always preferable from an asymptotic point-of-view. This result is in-line with a recent observation by Genest and Segers (2010) for the ordinary empirical copula process, where the rank-based estimator for the copula is more efficient for a broad class of positively associated copulas. For a similar observation regarding the analysis of censored data see, e.g., Portnoy (2010).
The remainder of this note is organized as following. In Section 2 we discuss the bias and the asymptotic variance of the two estimators $\tilde{\Lambda}_L$ and $\hat{\Lambda}_L$. While the bias is shown to be (almost) the same for both estimators the variance of $\hat{\Lambda}_L$ is shown to be substantially smaller. We investigate our findings for the example of a Clayton tail copula both theoretically and by means of a small simulation study.

2 Main result

In the subsequent developments we restrict ourselves to the investigation of lower tail dependence. Similar results hold in the upper tail. We are going to show that both estimators $\tilde{\Lambda}_L$ and $\hat{\Lambda}_L$ share a comparable bias under usual second order conditions, whereas the variance of $\hat{\Lambda}_L$ is substantially smaller than that of $\tilde{\Lambda}_L$. We begin with the discussion of the bias which can be derived from the decompositions

$$\sqrt{k}(\tilde{\Lambda}_L(x,y) - \Lambda_L(x,y)) = \sqrt{k}\left\{\frac{n}{k}\tilde{C}_n\left(\frac{kx}{n},\frac{ky}{n}\right) - \frac{n}{k}C\left(\frac{kx}{n},\frac{ky}{n}\right)\right\} + \sqrt{k}\left\{\frac{n}{k}C\left(\frac{kx}{n},\frac{ky}{n}\right) - \Lambda_L(x,y)\right\},$$
$$\sqrt{k}(\hat{\Lambda}_L(x,y) - \Lambda_L(x,y)) = \sqrt{k}\left\{\frac{n}{k}\hat{C}_n\left(\frac{kx}{n},\frac{ky}{n}\right) - \frac{n}{k}C\left(\frac{kx}{n},\frac{ky}{n}\right)\right\} + \sqrt{k}\left\{\frac{n}{k}C\left(\frac{kx}{n},\frac{ky}{n}\right) - \Lambda_L(x,y)\right\}. \tag{1.1}$$

The first summands in each line are the leading terms and converge weakly towards $\mathbb{B}_\Lambda$ and $\mathbb{G}_\Lambda$, respectively. Whereas the leading term in the first line in unbiased for every n, its counterpart in the second line is at least asymptotically unbiased. The summands on the right-hand side of the preceding decomposition are the same for both estimators and constitute the term which determines the asymptotic bias. Depending on the second order condition and on the limit behavior of $\sqrt{k}B(n/k)$ for $n \to \infty$ it may converge to 0 or to some function g, or its absolute value may blow up to ∞.
For these reasons an (asymptotic) comparison of the estimators $\tilde{\Lambda}_L$ and $\hat{\Lambda}_L$ must be based on a discussion of their asymptotic (co)variance. The following Theorem, in which we abbreviate Λ_L by Λ, is our main result.

Theorem 2.1. *Suppose that the partial derivatives of the tail copula $\dot{\Lambda}_1$ and $\dot{\Lambda}_2$ exist and are continuous on $(0,\infty)^2$. Then*

$$\operatorname{Cov}\{\mathbb{G}_\Lambda(x,y), \mathbb{G}_\Lambda(s,t)\} \le \operatorname{Cov}\{\mathbb{B}_\Lambda(x,y), \mathbb{B}_\Lambda(s,t)\}$$

for all $x, y, s, t \ge 0$. In particular, $\operatorname{Var}(\mathbb{G}_\Lambda(x,y)) \le \operatorname{Var}(\mathbb{B}_\Lambda(x,y))$.

Proof. The assertion is trivial if any of the variables equals zero, hence suppose $x, y, s, t > 0$. Multiplying out we have

$$\mathrm{Cov}\{\mathbb{G}_\Lambda(x,y), \mathbb{G}_\Lambda(s,t)\} - \mathrm{Cov}\{\mathbb{B}_\Lambda(x,y), \mathbb{B}_\Lambda(s,t)\} = \sum_{i=1}^{4} A_i - \sum_{i=5}^{8} A_i$$

where $A_i = A_i(x, y, s, t)$ is defined as

$$\begin{aligned}
A_1 &= \dot\Lambda_1(x,y)\dot\Lambda_1(s,t)(x \wedge s) & A_2 &= \dot\Lambda_1(x,y)\dot\Lambda_2(s,t)\Lambda(x,t) \\
A_3 &= \dot\Lambda_2(x,y)\dot\Lambda_1(s,t)\Lambda(s,y) & A_4 &= \dot\Lambda_2(x,y)\dot\Lambda_2(s,t)(y \wedge t) \\
A_5 &= \dot\Lambda_1(s,t)\Lambda(x \wedge s, y) & A_6 &= \dot\Lambda_2(s,t)\Lambda(x, y \wedge t) \\
A_7 &= \dot\Lambda_1(x,y)\Lambda(x \wedge s, t) & A_8 &= \dot\Lambda_2(x,y)\Lambda(s, y \wedge t).
\end{aligned}$$

For symmetry reasons we may suppose that $x \le s$. We will now prove that each summand A_i with $i = 1, \ldots, 4$ can be matched with a summand A_j, $j = 5, \ldots, 8$, such that $A_i - A_j \le 0$. This is done in the following way:

$$\begin{aligned}
A_1 - A_5 &= \dot\Lambda_1(s,t)\{\dot\Lambda_1(x,y)x - \Lambda(x,y)\} \\
A_2 - A_7 &= \dot\Lambda_1(x,y)\{\dot\Lambda_2(s,t)\Lambda(x,t) - \Lambda(x,t)\} \\
A_3 - A_8 &= \dot\Lambda_2(x,y)\{\dot\Lambda_1(s,t)\Lambda(s,y) - \Lambda(s, y \wedge t)\} \\
A_4 - A_6 &= \dot\Lambda_2(s,t)\{\dot\Lambda_2(x,y)(y \wedge t) - \Lambda(x, y \wedge t)\}.
\end{aligned}$$

Note that $\Lambda(x,y)/x = \Lambda(1, y/x)$ together with monotonicity of $\Lambda(1, \cdot)$ implies that the function $x \mapsto \Lambda(x,y)/x$ is non-increasing. Hence, for all $x, y > 0$

$$\dot\Lambda_1(x,y)x \le \Lambda(x,y), \qquad \dot\Lambda_2(x,y)y \le \Lambda(x,y), \tag{1.2}$$

where the second inequality follows analogously to the first one. This implies $A_1 - A_5 \le 0$. The assertion $A_2 - A_7 \le 0$ follows by boundedness of the partial derivative, i.e., $0 \le \dot\Lambda_2 \le 1$. Regarding the remaining two differences we distinguish two cases, namely $y \le t$ and $y > t$. In the former case, $A_3 - A_8 \le 0$ follows again from the upper bound 1 for $\dot\Lambda_1$, while $A_4 - A_6 \le 0$ follows from (1.2). In the case $y > t$, we can exploit $\Lambda(s, y) \le s$ and (1.2) to conclude on $A_3 - A_8 \le 0$. Regarding the last difference, we estimate

$$\dot\Lambda_2(x,y)t \le \Lambda(x,y)t/y = \Lambda(xt/y, t) \le \Lambda(x,t)$$

and hence $A_4 - A_6 \le 0$. The proof is finished. □

Remark 2.2.

a) In contrast to the copula-pendant, i.e., Proposition 1 in Genest and Segers (2010), Theorem 2.1 holds for every tail copula and not only for a subclass. An heuristic explanation for this observation is that tail-dependent random variables are strongly positively associated (at least in the tails). A copula being *left-tail decreasing* is positively associated as well, and this is exactly the class of copulas for which Genest and Segers (2010) showed the superiority of the rank-based copula estimator.

b) As a consequence of Theorem 2.1, by the same arguments as in Proposition 3 in Genest and Segers (2010), we can conclude that any real-valued estimator $\hat\theta = \Phi(\hat\Lambda_L)$ that is a non-decreasing and sufficiently smooth functional of $\hat\Lambda_L$ is preferable to the competitor $\tilde\theta = \Phi(\tilde\Lambda_L)$ from an asymptotic variance point-of-view. We omit the details of this observation.

Example 2.3. The lower tail copula arising from the bivariate Clayton copula is given by

$$\Lambda_L(x,y) = \lim_{t\to 0} t^{-1}\{(xt)^{-\theta} + (yt)^{-\theta} - 1\}^{-1/\theta} = \lim_{t\to 0}(x^{-\theta} + y^{-\theta} - t^{\theta})^{-1/\theta} = (x^{-\theta} + y^{-\theta})^{-1/\theta},$$

where $\theta > 0$. The first order partial derivatives are calculated as

$$\dot{\Lambda}_{L,1}(x,y) = (x^{-\theta} + y^{-\theta})^{-(1+\theta)/\theta} x^{-(1+\theta)}, \qquad \dot{\Lambda}_{L,2}(x,y) = (x^{-\theta} + y^{-\theta})^{-(1+\theta)/\theta} y^{-(1+\theta)}.$$

In Figure 1 we plot the graphs of the asymptotic variances of the estimators $\tilde{\Lambda}_L$ and $\hat{\Lambda}_L$ on the unit cube $[0,1]^2$ for $\theta = 1$ corresponding to a coefficient of tail dependence of 0.5. The difference is seen to be substantial, especially close to the axis.

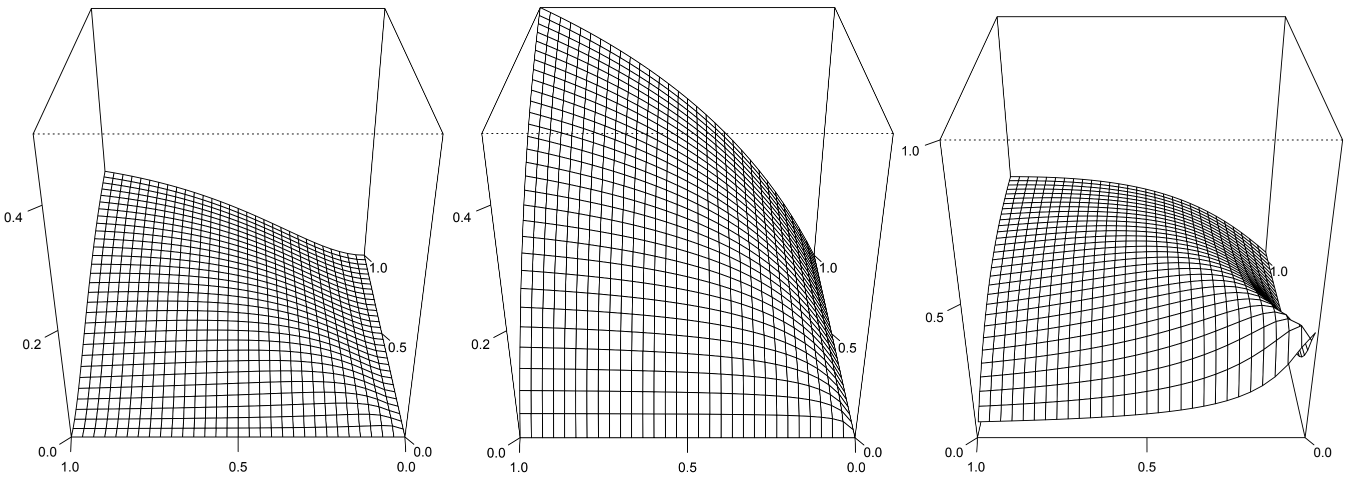

Figure 1: The graphs of $\mathrm{Var}(\mathbb{G}(x,y))$ (left, unknown marginals), of $\mathrm{Var}(\mathbb{B}(x,y))$ (middle, known marginals) and of the relative efficiency $\mathrm{Var}(\mathbb{G}(x,y))/\mathrm{Var}(\mathbb{B}(x,y))$ (right) for the Clayton tail copula with parameter $\theta = 1$ and for $(x,y) \in [0,1]^2$.

In order to investigate the influence of the parameter θ we restrict ourselves to the estimation of the coefficient of lower tail dependence. The calculations above show that $\lambda_L = \Lambda_L(1,1) = 2^{-1/\theta}$ for the Clayton tail copula. The competitive estimators are $\tilde{\lambda}_L = \tilde{\Lambda}_L(1,1)$ and $\hat{\lambda}_L = \hat{\Lambda}_L(1,1)$, and a careful calculation reveals that their asymptotic variances are given by

$$\mathrm{Var}(\mathbb{B}_\Lambda(1,1)) = 2^{-1/\theta}, \qquad \mathrm{Var}(\mathbb{G}_\Lambda(1,1)) = 2^{-1/\theta}\{1 + 2^{-1/\theta-1} + 2^{-2/\theta-1} - 2^{-1/\theta+1}\}.$$

In Figure 2, these variances are plotted as a function of θ. It can be seen that the rank-based estimator becomes substantially better with increasing degree of tail dependence.

Finally, we investigate the finite-sample performance of the estimators for λ_L by means of a small simulation study. To this end, we simulate i.i.d. samples of size $n = 1.000$ of the Clayton tail copula with parameter $\theta = 0.5$ such that the coefficient of tail dependence is $\lambda_L = 0.25$. Our objective is the estimation of λ_L and we investigate both the squared bias and the asymptotic variance as a function of the parameter k. As usual in extreme value theory, larger values of k result in a larger bias, whereas the variance decreases in k. The results, which are based on 10.000 repetitions, are plotted in Figure 3. We clearly see the expected superiority of $\hat{\lambda}_L$ in Variance and Mean Squared Error. The bias of both estimators is indeed comparable as indicated by decomposition (1.1) at the beginning of this Section.

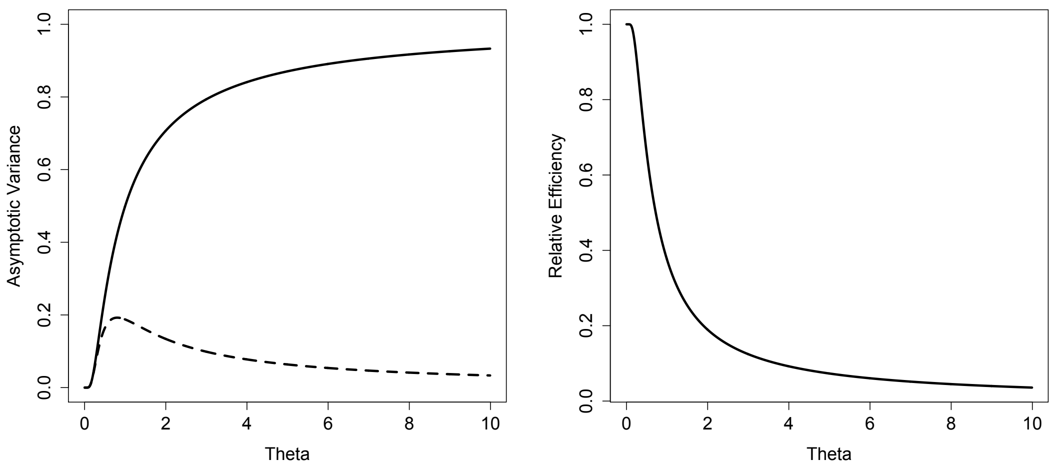

Figure 2: Left: Asymptotic Variance $\mathrm{Var}(\mathbb{G}(1,1))$ (dashed line) and $\mathrm{Var}(\mathbb{B}(1,1))$ (solid line) for the Clayton tail copula as a function of θ. Right: Asymptotic relative efficiency $\mathrm{Var}(\mathbb{G}(1,1))/\,\mathrm{Var}(\mathbb{B}(1,1))$

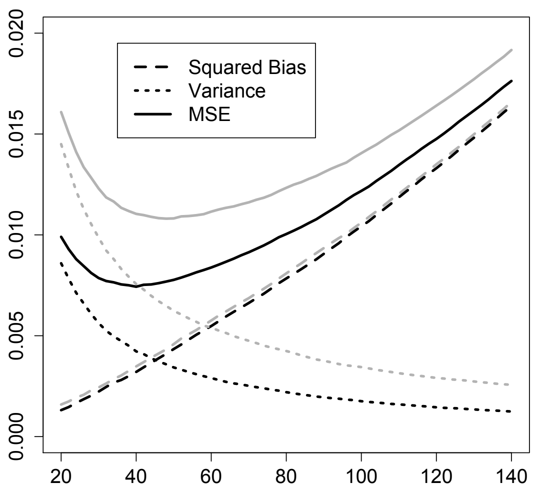

Figure 3: Squared Bias (dashed lines), Variance (dotted lines) and MSE (solid lines) of the estimators $\hat{\Lambda}_L(1,1)$ (black lines) and $\tilde{\Lambda}_L(1,1)$ (gray lines) for $\lambda_L = 0.25$ in the Clayton tail copula model, as a function of the parameter k.

Note also that the difference in variance entails a different "optimal" choice of k for the two estimators, which is seen to be slightly larger for the estimator based on known marginals. This discrepancy reveals that there is no perfect global answer to the question of *where the tail begins*.

Acknowledgements. This work has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG). The author is grateful to Stanislav Volgushev for discussions and suggestions concerning this manuscript.

References

Bücher, A. and Dette, H. (2011). Multiplier bootstrap approximations of tail copulas – with applications. *arXiv:1102.0110.* to appear in Bernoulli.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimates of the stable tail dependence functions. *J. Multivar. Anal.*, 64:25–47.

Einmahl, J. H. J., Krajina, A., and Segers, J. (2012). An m-estimator for tail dependence in arbitrary dimensions. *arXiv:1112.0905.* to appear in Annals of Statistics.

Genest, C. and Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.*, 101(8):1837–1845.

Huang, X. (1992). *Statistics of bivariate extreme values.* PhD thesis, Tinbergen Institute Research Series, Netherlands.

Pickands, J. (1981). Multivariate extreme value distributions. *Bull. Inst. Internat. Statist.*, 49(2):859–878, 894–902.

Portnoy, S. (2010). Is ignorance bliss: fixed vs. random censoring. In *Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jurečková*, volume 7 of *Inst. Math. Stat. Collect.*, pages 215–223. Inst. Math. Statist., Beachwood, OH.

Schmidt, R. and Stadtmüller, U. (2006). Nonparametric estimation of tail dependence. *Scand. J. Statist.*, 33:307–335.